STA 2201/442 Assignment 11

1. Consider a two-factor analysis of variance in which each factor has two levels. Use this regression model for the problem:

$$Y_i = \beta_0 + \beta_1 d_{i,1} + \beta_2 d_{i,2} + \beta_3 d_{i,1} d_{i,2} + \epsilon_i,$$

    where $d_{i,1}$ and $d_{i,2}$ are dummy variables.

    (a) Make a two-by-two table showing the four treatment means in terms of β values. Use *effect coding*. In terms of the β values, state the null hypothesis you would use to test for

        i. Main effect of the first factor
        ii. Main effect of the second factor
        iii. Interaction

    (b) Make a two-by-two table showing the four treatment means in terms of β values. Use *indicator dummy variables* (zeros and ones). In terms of the β values, state the null hypothesis you would use to test for

        i. Main effect of the first factor
        ii. Main effect of the second factor
        iii. Interaction

    (c) Which dummy variable scheme do you like more?

2. I know this is pretty gruesome, but the data are real. An experiment in dentistry seeks to test the effectiveness of a drug (HEBP) that is supposed to help dental implants become more firmly attached to the jaw bone. This is an initial test on animals. False teeth were implanted into the leg bones of rabbits, and the rabbits were randomly assigned to receive either the drug or a saline solution (placebo). Technicians administering the drug were blind to experimental condition.

    Rabbits were also randomly assigned to be "sacrificed" after either 3, 6, 9 or 12 days. At that time, the implants were pulled out of the bone by a machine that measures force in newtons and stiffness in newtons/mm. For both of these measurements, higher values indicate more healing. A measure of "pre-load stiffness" in newtons/mm is also available for each animal. This may be another indicator of how firmly the false tooth was implanted into the bone, but it might even be a covariate. Nobody can seem to remember what "preload" means, so we'll ignore this variable for now.

The data are available in the file `bunnies.data`. The variables are

- Identification code
- Time (3,6,9,12 days of healing)
- Drug (1=HEBP, 0=saline solution)
- Stiffness in newtons/mm
- Force in newtons
- Preload stiffness in newtons/mm

Classify the factors as within cases or between cases. Then please do the following using SAS. Remember that in the `input` statement, character-valued variables should be followed by a dollar sign.

(a) Use `proc freq` to find out how many rabbits are in each experimental condition.

(b) Using `proc glm`, conduct a two-way ANOVA, with force as the dependent variable. Use the means statement to get cell means and marginal means. Be prepared to answer the following questions about each of the significance tests that SAS produces by default (I count 4 default tests).

    i. What is the value of the test statistic? The answer is a number from your printout.

    ii. What proportion of the remaining variation is explained? You can use `proc iml`, or just be ready to do it with a calculator. The formula will be supplied with the quiz if this question is asked.

    iii. What is the p-value? The answer is a number from your printout.

    iv. Do you reject the null hypothesis at the 0.05 level? Yes or No.

    v. What, if anything, do you conclude? This is not the place for statistical jargon. "What do you conclude" means say something about the drug, healing, time – something like that.

(c) I know this is a bit redundant with the preceding question, but *did the drug work?* If the results justify an answer, then answer Yes or No.

(d) Now, make a table with a row for each treatment combination. Give the coefficients of the constrast or contrasts that would be used to test for

    i. The main effect of Drug

    ii. The main effect(s) of Time

    iii. The Drug by Time interaction.

(e) Make another table with a row for each treatment combination. Make columns showing the dummy variables for effect coding.

(f) Give $E[Y|\mathbf{X} = \mathbf{x}]$ for a regression model with both main effects and the interaction. Use your variable names from the preceding question.

(g) In terms of the β values of your regression model, give the null hypothesis you would test in order to answer each of the following questions.

    i. Averaging across time periods, is there a differnece between the drug and placebo in mean force required to extract the tooth?

    ii. Averaging across drug and placebo, is does elapsed time affect the mean force required to extract the tooth?

    iii. Does the effect of the drug depend upon elapsed time?

(h) Now please return to SAS. Using `proc reg` and cell means coding, conduct tests to answer the following questions. Just do regular one-at-a-time (custom) tests. Don't bother with any Bonferroni correction this time. Just consider one dependent variable: Force. As usual, we are guided by the $\alpha = 0.05$ significance level.

    i. Are the marginal means different at 3 and 6 days?

    ii. Are the marginal means different at 6 and 9 days?

    iii. Are the marginal means different at 9 and 12 days?

    iv. Is there a difference between Drug and Placebo just at 3 days?

    v. Is there a difference between Drug and Placebo just at 6 days?

    vi. Is there a difference between Drug and Placebo just at 9 days?

    vii. Is there a difference between Drug and Placebo just at 12 days?

    viii. Be able to answer questions like these for each test:

        A. What is the value of the test statistic? The answer is a number from your printout.

        B. What proportion of the remaining variation is explained? You can use `proc iml`, or just be ready to do it with a calculator. The formula will be supplied with the quiz if this question is asked.

        C. What is the p-value? The answer is a number from your printout.

        D. Do you reject the null hypothesis at the 0.05 level? Yes or No.

        E. What, if anything, do you conclude? This is not the place for statistical jargon. "What do you conclude" means say something about the drug, healing, time – something like that.

Please bring your `SAS` log file and list file to the quiz. Remember, the log file is not the same as a plain listing of the program. Please be sure that the log file and list file come from the same SAS run. *Do not write anything on your printouts in advance except maybe your name* (and student number, if you wish).